

**This Page Is Inserted by IFW Operations
and is not a part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- **BLACK BORDERS**
- **TEXT CUT OFF AT TOP, BOTTOM OR SIDES**
- **FADED TEXT**
- **ILLEGIBLE TEXT**
- **SKEWED/SLANTED IMAGES**
- **COLORED PHOTOS**
- **BLACK OR VERY BLACK AND WHITE DARK PHOTOS**
- **GRAY SCALE DOCUMENTS**

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problems Mailbox.**

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2000 Japan Patent Office

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平9-311805

(43)公開日 平成9年(1997)12月2日

(51)Int.Cl. ⁶	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 12/00	5 2 0		G 0 6 F 12/00	5 2 0 P
	5 1 0			5 1 0 A
17/21			15/20	5 7 0 D

審査請求 未請求 請求項の数20 O L (全 14 頁)

(21)出願番号 特願平8-129899

(22)出願日 平成8年(1996)5月24日

(71)出願人 000001007

キヤノン株式会社

東京都大田区下丸子3丁目30番2号

(72)発明者 伊藤 史朗

東京都大田区下丸子3丁目30番2号キヤノン株式会社内

(72)発明者 柴田 昇吾

東京都大田区下丸子3丁目30番2号キヤノン株式会社内

(72)発明者 上田 隆也

東京都大田区下丸子3丁目30番2号キヤノン株式会社内

(74)代理人 弁理士 丸島 儀一

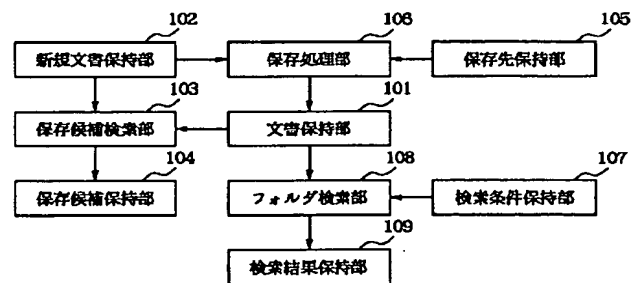
最終頁に続く

(54)【発明の名称】 文書処理方法及び装置

(57)【要約】

【課題】 新規文書を保持するのに適したファイルの選択作業及び所望の検索条件に一致する文書を探すのに適したフォルダの選択作業を容易にする。

【解決手段】 新規文書を保存するのに適したフォルダを検索する保存する保存候補検索部103を有し、ここで検索された保存候補のフォルダをユーザに報知する。また、検索条件に一致するフォルダを検索するフォルダ検索部108を有し、ここで検索されたフォルダをユーザに報知する。



【特許請求の範囲】

【請求項 1】 フォルダ及びフォルダに属する文書を保持する文書保持手段と、
新規文書に適した保存先フォルダを、該新規文書を解析することにより前記文書保持手段に保持されているフォルダより判定する保存先判定手段と、
前記保存先判定手段により判定された保存先フォルダを報知する報知手段とを有することを特徴とする文書処理装置。

【請求項 2】 フォルダ及びフォルダに属する文書を保持する文書保持手段と、
新規文書を保持する新規文書保持手段と、
前記新規文書保持手段に保持されている文書の保存先として適するフォルダを検索する保存候補検索手段と、
前記保存候補検索手段により検索された保存先の候補を出力する保存候補出力手段と、
前記新規文書保持手段に保持されている新規文書の保存先を選択する保存先選択手段と、
前記新規文書保持手段に保持されている文書を前記保存先選択手段により選択された保存先のフォルダに属するよう前記文書保持手段に保持する保存処理手段とを有する事を特長とする文書処理装置。

【請求項 3】 フォルダ及びフォルダに属する文書を保持する文書保持手段と、
新規文書を保持する新規文書保持手段と、
前記新規文書保持手段に保持されている文書の保存先として適するフォルダ毎に前記新規文書を分類する文書分類手段と、
前記文書分類手段により分類された結果を出力する分類結果出力手段と、
前記文書分類手段により分類された結果の変更を支持する保存先変更手段と、
前記新規文書保持手段に保持されている文書を前記保存先変更手段により変更された保存先のフォルダに属するよう前記文書保持手段に保持する保存手段とを有することを特徴とする文書処理装置。

【請求項 4】 前記保存先変更手段は、前記分類結果出力手段により出力された分類結果を削除することにより変更を指示することを特徴とする請求項 3 に記載の文書処理装置。

【請求項 5】 前記文書保持手段に保持する新しいフォルダを作成するフォルダ作成手段を有することを特徴とする請求項 3 に記載の文書処理装置。

【請求項 6】 前記フォルダは、当該フォルダに属して保持されている文書の文書特徴の代表値の情報を含むことを特徴とする請求項 3 に記載の文書処理装置。

【請求項 7】 前記文書特徴の代表値は、前記フォルダに属して保持されている文書の文書特徴の平均値とすることを特徴とする請求項 6 に記載の文書処理装置。

【請求項 8】 前記新規文書の文書特徴と、前記フォル

ダに含まれる文書特徴の代表値とを比較し、類似していると判定されるフォルダを該新規文書の保存先として適すると判定することを特徴とする請求項 6 に記載の文書処理装置。

【請求項 9】 前記文書の保存先として適するフォルダを複数判定し、前記複数判定されたフォルダに関して前記分類を出力することを特徴とする請求項 3 に記載の文書処理装置。

【請求項 10】 前記分類結果出力手段は表示器とすることを特徴とする請求項 3 に記載の文書処理装置。

【請求項 11】 検索条件を保持する検索条件保持手段と、
前記検索条件保持手段に保持されている検索条件に従って、前記文書保持手段に保持されているフォルダを検索するフォルダ検索手段と、
前記フォルダ検索手段により検索されたフォルダを報知する検索結果報知手段とを有することを特徴とする請求項 3 に記載の文書処理装置。

【請求項 12】 フォルダ及びフォルダに属する文書を保持する文書保持手段と、
検索条件を保持する検索条件保持手段と、
前記検索条件保持手段に保持されている検索条件に従って、前記文書保持手段に保持されているフォルダを検索するフォルダ検索手段と、
前記フォルダ検索手段により検索されたフォルダを報知する検索結果報知手段とを有することを特徴とする文書処理装置。

【請求項 13】 前記フォルダ検索手段は、前記保持された検索条件に一致する文書を保持しているフォルダを検索することを特徴とする請求項 12 に記載の文書処理装置。

【請求項 14】 前記検索条件は文字列とすることを特徴とする請求項 12 に記載の文書処理装置。

【請求項 15】 前記検索結果報知手段は、フォルダの名称を報知することを特徴とする請求項 12 に記載の文書処理装置。

【請求項 16】 前記検索結果報知手段は前記検索条件との一致スコアが高い順に検索結果を報知することを特徴とする請求項 12 に記載の文書処理装置。

【請求項 17】 フォルダ及びフォルダに属する文書を保持する文書保持手段を利用し、
新規文書に適した保存先フォルダを、該新規文書を解析することにより前記文書保持手段に保持されているフォルダより判定し、
前記判定された保存先フォルダを報知することを特徴とする文書処理方法。

【請求項 18】 フォルダ及びフォルダに属する文書を保持する文書保持手段を利用し、
新規文書を保持し、
前記保持されている文書の保存先として適するフォルダ

を検索し、
前記検索された保存先の候補を出力し、
前記保持されている新規文書の保存先を選択し、
前記保持されている文書を前記選択された保存先のフォルダに属するよう前記文書保持手段に保持することを特徴とする文書処理方法。

【請求項19】 フォルダ及びフォルダに属する文書を保持する文書保持手段を利用し、
新規文書を保持し、
前記保持されている文書の保存先として適するフォルダ毎に前記新規文書を分類し、
前記分類された結果を出力し、
前記分類された結果の変更を指示し、
前記保持されている新規文書を前記変更された保存先のフォルダに属するよう前記文書保持手段に保持することを特徴とする文書処理方法。

【請求項20】 フォルダ及びフォルダに属する文書を保持する文書保持手段を利用し、
検索条件を保持し、
前記保持されている検索条件に従って、前記文書保持手段に保持されているフォルダを検索し、
前記検索されたフォルダを報知することを特徴とする文書処理方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、例えば電子新聞、電子出版、電子回覧などで用いられる電子化文書の収集と収集した文書の管理を行なう文書処理装置に関するものである。

【0002】

【従来の技術】 近年、電子新聞、電子出版、電子回覧といった電子化文書による情報流通が広まってきている。こうした文書を収集して管理するために文書処理装置が用いられる。

【0003】 従来この種の文書処理装置では、新たに到着した文書を列挙し、それらを利用者が閲覧して必要な文書を収集する方法が採られている。収集した文書の保存先としてはフォルダが用意され、利用者は列挙されているフォルダから、保存先を選び、当該文書を保存する方法が採られている。保存した文書を利用する場合には、列挙されている中からフォルダを選び、フォルダ中の文書を利用する方法が採られている。ここで、フォルダを階層的に保持して、階層を辿ってフォルダを選べるようにする方法が採られることもある。

【0004】 この文書処理装置を利用する場合、利用者は自らの視点からみて関連する文書を同一のフォルダに入れて保存し、利用時に自らの視点に沿ってフォルダを選択し、所望の文書を利用することになる。

【0005】 フォルダを利用せずに収集した文書を管理する文書処理装置としては、文書の属性を用いて検索を

行なうデータベースマネジメントシステムや文書にキーワードを付与して検索を行なう情報検索システム、あるいは文書の本文中の検索語の有無を調べて検索を行なう全文検索システムなどがある。

【0006】

【発明が解決する課題】 しかしながら、上記従来例の装置では、収集時に保存先のフォルダを探すことや、利用時に適切なフォルダを探すことが困難となり文書の収集および利用の効率が低下するという問題点がある。こうした現象は、一つにはフォルダが多数になる場合に起きる、フォルダが多数になった場合、列挙されたフォルダを一覧して適切なフォルダを探し出すことは困難になる、フォルダを階層的に保持することで、この問題点はある程度緩和されるが、利用者の視点は時間推移と共に変化することが多く、以前に作成した階層構造がその後の視点に合わなくなると、階層を辿ることが難しくなり、フォルダを探し出すことが困難になる、あるいは、あるフォルダを最後に利用してから時間が相当に経過している場合には、そのフォルダが何であったか、あるいはそのフォルダの存在そのものを利用者が忘れていることが多い、こうした場合も、フォルダを探すことが困難になる。フォルダを探すことが困難になると、収集時には、フォルダ数を少なく抑えたり、不適当なフォルダに入れたり、複数のフォルダに保存することを減らしたり、保存を止めたりすることになる。こうした場合、フォルダが視点を正しく反映しなくなり、利用時にフォルダから所望の文書を探すのが困難になるという問題点も生じる。

【0007】 一方、データベースマネジメントシステムや情報検索システムなどを使って文書を管理しようとすると、文書の収集時に属性値やキーワードなどを付与する必要があり、収集時の負荷が高い。この種の文書処理装置は、個人が日常的に用いることが多いので、収集時の負荷が高いことは問題である。

【0008】 文書に予め付与されている属性やキーワードだけを用いてデータベースマネジメントシステムや情報検索システムを利用したり、全文検索システムを利用して文書を管理する方法では、利用者の視点に沿った検索が困難になるという問題点がある。

【0009】 本発明は上記の問題に鑑みてなされたものであり、利用者の視点に沿って文書を管理するとともに、文書の収集と利用を容易に行なうことを可能とする文書処理方法及び装置を提供することを目的とする。

【0010】

【課題を解決するための手段】 上記課題を解決するために、本発明は好ましくはフォルダ及びフォルダに属する文書を保持し、新規文書に適した保持先フォルダを、該新規文書を解析することにより前記文書保持手段に保持されているフォルダより判定し、前記判定された保存先フォルダを報知する文書処理方法及び装置を提供する。

【0011】上記課題を解決する為に、本発明はフォルダ及びフォルダに属する文書を保持し、新規文書を保持し、前記保持されている文書の保存先として適するフォルダを検索し、前記検索された保存先の候補を出力し、前記保持されている新規文書の保存先を選択し、前記新規文書を前記選択された保存先のフォルダに属するよう保持する文書処理方法及び装置を提供する。

【0012】上記課題を解決する為に、本発明はフォルダ及びフォルダに属する文書を保持し、新規文書を保持し、前記保持されている文書の保存先として適するフォルダ毎に前記新規文書を分類し、前記分類された結果を出力し、前記分類された結果の変更を指示し、前記新規文書を前記変更された保存先のフォルダに属するよう保持する文書処理方法及び装置を提供する。

【0013】上記課題を解決する為に、本発明は好ましくは前記分類結果出力手段により出力された分類結果を削除することにより前記保存先の変更を指示する。

【0014】上記課題を解決する為に、本発明は好ましくは新しいフォルダを作成する。

【0015】上記課題を解決する為に、本発明は好ましくは前記フォルダは、当該フォルダに属して保持されている文書の文書特徴の代表値の情報を含む。

【0016】上記課題を解決する為に、本発明は好ましくは前記文書特徴の代表値は、前記フォルダに属して保持されている文書の文書特徴の平均値とする。

【0017】上記課題を解決する為に、本発明は好ましくは前記新規文書の文書特徴と、前記フォルダに含まれる文書特徴の代表値とを比較し、類似していると判定されるフォルダを該新規文書の保存先として適すると判定する。

【0018】上記課題を解決する為に、本発明は好ましくは前記文書の保存先として適するフォルダを複数判定し、前記複数判定されたフォルダに関して前記分類を出力する。

【0019】上記課題を解決する為に、本発明は好ましくは前記分類結果出力手段は表示器とする。

【0020】上記課題を解決する為に、本発明は好ましくは検索条件を保持し、前記保持されている検索条件に従って、前記保持されているフォルダを検索し、前記検索されたフォルダを報知する。

【0021】上記課題を解決する為に、本発明は好ましくは前記フォルダの検索は、前記保持された検索条件に一致する文書を保持しているフォルダを検索する。

【0022】上記課題を解決する為に、本発明は好ましくは前記検索条件は文字列とする。

【0023】上記課題を解決する為に、本発明は好ましくは前記検索結果は、検索されたフォルダの名称を報知する。

【0024】上記課題を解決する為に、本発明は好ましくは前記検索結果の報知は前記検索条件との一致スコア

が高い順に行なう。

【0025】

【発明の実施の形態】以下、図面を参照して本発明の実施の形態を詳細に説明する。

【0026】図1は、本発明の一実施の形態に係る装置の基本構成を示すブロック図である。

【0027】同図において、101は、フォルダ及びフォルダに属する文書を保持する文書保持部である。102は、新たに到着した文書を保持する新規文書保持部である。103は新規文書保持部102に保持されている文書の保存先として適するフォルダを検索する保存候補検索部である。104は、保存候補検索部103により検索された保存先の候補を保持する保存候補保持部である。105は、保存候補保持部104に保持されている保存候補の中から利用者により選択された保存先を保持する保存先保持部である。106は、新規文書保持部102に保持されている文書を保存先保持部105に保持されている保存先のフォルダに属するよう文書保持部101に保持する処理を行なう保存処理部である。107は、フォルダに対する検索条件を保持する検索条件保持部である。108は、検索条件保持部107に保持されている検索条件に従って、文書保持部101に保持されているフォルダを検索するフォルダ検索部である。109は、フォルダ検索部108により検索されたフォルダを保持する検索結果保持部である。

【0028】図2は、本実施例の文書処理装置の具体的な構成を示す図である。同図において、201はCPUであり、ROM203に保持されたプログラムに従って動作する。

【0029】202はROMであり、新規文書保持部102と保存候補保持部104と保存先保持部105と検索条件保持部107と検索結果保持部と上記プログラムの動作に必要な記憶領域とを提供する。203はROMであり、後述するフローチャートの手順を実現するプログラムを保持する。204はディスク装置であり、文書保持部101を実現する。205はバスである。206は文字や画像等を表示し得る表示器であり、CRTや液晶表示器等である。207はキーボードやポインティングデバイス等の入力手段である。

【0030】本実施例では、文書保持部101は、文書のリストとフォルダのリストを保持する構造になっている。文書dは次のようなデータから構成される。

【0031】 $d = (t, v(d))$

ここで、tは文書のテキストデータであり、v(d)はベクトル空間モデルに基づきテキストtの特徴を表現したベクトルである。また、フォルダfは次のようなデータから構成される。

【0032】 $f = (1, D, v(f))$

ここで、1はラベルであり、利用者がフォルダを視認するための文字列であって、利用者がキーボード等の入力

手段207により入力したものや、或いは自動的に付加されたものであっても良い。Dは、フォルダに保存される文書の集合であり、Dは空でもよい。v f は、そのフォルダfに格納されている全ての文書dについてのベクトル $v(d)$ ($d \in D$)の平均をとったベクトルである。以下、文書保持部101に保持されているフォルダの数をNで表す。新規文書保持部102は、一つの文書を保持する構造になっている。保存候補保持部104及び保存先保持部105及び検索結果保持部109は、フォルダの番号を保持するリストとなっている。検索条件保持部107は、検索語とその論理関係を示す検索式を保持する構造になっている。

【0033】以下、図3のフローチャートを参照して、本実施の形態の文書処理装置における保存候補検索処理の手順を説明する。

【0034】ステップS301では、新規文書保持部102に新たに到着した文書のテキストt(n)が保持されたか否かを調べ、文書が保持された場合はステップS302に移る。保持されていない場合は、ステップS301を繰り返し、新たな文書のテキストt(n)が新規文書保持部102に保持されるのを待つ。ここで新たな文書のテキストt(n)は、ユーザの入力指示のタイミング、或いはテキスト供給元からの自動供給のタイミングで新規文書保持部102に到着する。

【0035】ステップS302では、テキストt(n)に対して特徴ベクトル $v(d_n)$ を作成し、t(n)と共に新規文書保持部102に保持する。そして、ステップS303に移る。

【0036】ステップS303では文書保持部101に保持されているフォルダの情報に順次アクセスするためのフォルダ番号のカウンタxの値を初期値の1に設定する。そして、ステップS304に移る。

【0037】ステップS304では、S305～S307の処理が文書保持部101に保持されている全てのフォルダに対して行なわれたか判定すべく、カウンタxの値と文書保持部101中のフォルダの数Nを比較し、 $x \leq N$ であればステップS305に移る。 $x > N$ であれば図3のフローチャートに示す保存候補検索処理を終了する。

【0038】ステップS305では、文書保持部101中のx番目のフォルダf(x)と新規文書d(n)に対してスコア $s = g(v(d_n), v(fx))$ を計算する。ここで、関数gは以下に示すものである。

【0039】 $g(v(1), v(2)) = (v(1) \cdot v(2)) / (|v(1)| |v(2)|)$
スコアSを求めた後ステップS306に移る。

【0040】ステップS306では、ステップS305に求めたスコアsの昇順になるよう保存候補保持部104にS305でスコアを求めたフォルダの番号xをスコアSとともに保持する。そして、ステップS307に移

る。

【0041】ステップS307では、カウンタxの値を1増やす。そして、ステップS304に戻る。

【0042】図3のフローチャートにより説明した保存候補検索処理により求められ、保存候補保持部104に保持された保存先の候補フォルダに関する情報(フォルダのラベル等)は、表示器206に保持された順(スコアSの昇順)に新規文書保持部102に保持されている文書を対応づけて表示し、ユーザに報知される。

【0043】尚、この表示器に表示するフォルダは、保存候補保持部104に保持された全フォルダでも良いし、或いはスコアSや個数Mにより選択される上位のフォルダのみでも良い。

【0044】次に、図4のフローチャートを参照して、本実施の形態の文書処理装置における文書の保存処理の手順を説明する。

【0045】ステップS401では、保存先保持部105に保存先フォルダのリストFが保持されているか否かを調べ、保持されている場合はステップS402に移る。保持されていない場合は、ステップS401を繰り返し、保存先保持部105にリストFが保持されるのを待つ。このリストFは、ユーザによりキーボード等の入力手段によって入力されるフォルダ列であり、ユーザは保存候補保持部104に保持されたデータに基づいて報知される保存先の候補フォルダ情報を考慮した上でリストFを入力する。

【0046】ステップS402では、その時点でアクセスするフォルダがリストFの何番目のフォルダであるかを示す。カウンタxの値を初期値の1に設定する。そして、ステップS403に移る。

【0047】ステップS403では、カウンタxの値と保存先の個数|F|を比較し、 $x \leq |F|$ であればステップS404に移る。 $x > |F|$ であれば図4のフローチャートに示す文書の保存処理を終了する。

【0048】ステップS404では、保存先保持部105中のx番目に対応するフォルダf(Fx)の文書リストD(Fx)に新規文書d(n)を加える。さらに、このd(n)が加えられた新たなD(Fx)に対して、ベクトル $v(d)$ ($d \in D(Fx)$)の平均を計算し、新たな $v(f(Fx))$ とする。そして、ステップS405に移る。

【0049】ステップS405では、カウンタxの値を1増やす。そして、ステップS403に戻る。

【0050】次に、図5のフローチャートを参照して、本実施の形態の文書処理装置におけるフォルダ検索処理の手順を説明する。

【0051】ステップS501では、検索条件保持部107に検索条件cが保持されたか否かを調べ、保持された場合はステップS502に移る。保持されていない場合は、ステップS501を繰り返し、検索条件cが、保

持されるのを待つ。この検索条件cは、ユーザによりキーボード等の入力手段から入力される語或いは文章の文字列である。

【0052】ステップS502では、その時点でアクセスするフォルダが文書保持部101に保持されている全フォルダの中で何番目のフォルダであるかを示すカウンタxの値をデフォルトの1に設定する。そして、ステップS503に移る。

【0053】ステップS503では、カウンタxの値と文書保持部101中のフォルダの全数Nを比較し、 $x \leq N$ であればステップS504に移る。 $x > N$ であれば図5のフローチャートに示すフォルダ検索処理を終了する。

【0054】ステップS504では、文書保持部101中のx番目のフォルダf(x)と検索条件cに対してスコアsを次の式により計算する。

【0055】

【外1】

$$s = \frac{\sum_{d=1}^{100} f(c,d)}{|D(x)|}$$

【0056】なお、関数fは、パターンマッチングにより文書中に検索語cがあるか否かを調べ、cが文書中にある場合にはその文書については $f(c, d) = 1$ とし、Cが文書中にある場合は $f(c, d) = 0$ とするものである。この判断をx番目のフォルダ内の全文書D(x)について行う。従って、上述のスコアsは、x番目のフォルダ内に、検索語cがある文書がいくつあって、全体の文書数 $|D(x)|$ で割ることによって割合を求めたものである。

【0057】S505でx番目のフォルダについての検索スコアSを求めた後、ステップS505に移る。

【0058】ステップS505では、ステップS504で求めたスコアsの昇順になるよう検索結果保持部109にS504でスコアSを求めたフォルダの番号xをスコアSとともに保持する。そして、ステップS506に移る。

【0059】ステップS506では、カウンタxの値を1増やす。そして、ステップS503に戻る。

【0060】図5のフローチャートにより説明したフォルダ検索処理により求められ、検索結果保持部109に保持された検索結果のフォルダに関する情報（フォルダのラベル等）は、表示器206に保持された順（スコアsの昇順）に検索語cと対応づけて表示し、ユーザに報知される。尚、この表示器206に表示されるフォルダは、検索結果保持部109に保持された全フォルダでも良いし、或いはスコアSや個数Mにより選択される上位フォルダのみでも良い。

【0061】本実施の形態の文書処理装置を用いると、文書の収集時においては、新規文書の保存先として最適なフォルダが保存候補保持部104の先頭に保持され

る。

【0062】そこで、例えば、保存候補保持部104に保持されている順に従ってフォルダのラベルを表示することにより利用者は表示の先頭周辺を見ることで、保存先を容易に選ぶことができる。文書の利用時には、検索条件を指定すると、その検索条件に良く合う文書が集まるフォルダを検索することができるので、文書を探すフォルダを絞り込むことが可能となり、検索処理を効率的に行える。

【0063】以上のように本実施例の文書処理装置を用いることで、視点に沿って文書をフォルダに保持すると共に、文書の収集と利用が容易になる。

【0064】尚、本実施の形態においては、文書の収集と検索利用の双方における処理を容易にする機能を実現する場合について説明したが、本発明はこれに限定されるものではない。収集もしくは利用のみを容易にする機能を実現したものでもよい。この場合の実施例のブロック図を図6と図7に示す。なお、前述の図1の構成と比較すると明らかなように、図6の構成601から607は、図1の101から107に対応している。また、図7の構成701から704は、図1の101、107、108、109に対応している。

【0065】（他の実施例1）上記実施例においては、一つの文書ごとに保存先のフォルダの候補を検索して当該文書に対する保存先のフォルダの候補を提示することで収集に要する作業を容易ならしめる場合について説明したが、本発明はこれに限定されるものではない。本実施の形態では、新たに到着する文書の保存先として適していると判断されるフォルダごとに分類し、フォルダごとに対応した文書を提示することで収集に要する作業を容易にする例について説明する。

【0066】図8は、本実施の形態に係る装置の基本構成を示すブロック図である。同図において、801は、フォルダに及びフォルダに属する文書を保持する文書保持部である。802は、新たに到着した文書を保持する新規文書保持部である。803は、新規文書保持部802に保持されている文書の保存先として適するフォルダに文書を分類する文書分類部である。804は、文書分類部803により分類された結果を保持する分類結果保持部である。805は、保存処理の対象とする文書を保持する保存対象文書保持部である。806は、分類結果保持部804に保持されている分類結果から保存対象文書保持部805に保持されている保存対象文書の保存先を作成する保存先作成部である。807は、保存先作成部806により作成された保存先を保持する保存先保持部である。808は、利用者により保存先保持部807に保持されている保存先を変更する保存先変更部である。809は、保存対象文書保持部805に保持されている文書を保存先保持部807に保持されている保存先のフォルダに属するよう文書保持部801に保持する処

理を行なう保存処理部である。

【0067】本実施の形態では、分類結果保持部804は、フォルダfごとに分類された文書の集合を保持する構造になっている。保存対象文書保持部は、保存前の文書の一つを保持する構造になっている。文書保持部801、新規文書保持部802、保存先保持部807は、先に説明した図1の構成101、102、105と同様である。

【0068】本実施例における装置構成は先に説明した図2と同様であり、ここでは説明を省略する。

【0069】以下、図9のフローチャートを参照して、本実施の形態の文書処理装置における文書分類処理の手順を説明する。

【0070】ステップS901では、新規文書保持部802に新たに到着した文書のテキストt(n)が保持されたか否かを調べ、文書が保持された場合はステップS902に移る。保持されていない場合は、ステップS901を繰り返し、新たに到着した文書のテキストt(n)が保持されるのを待つ。

【0071】ステップS902では、テキストt(n)に対して特徴ベクトルv(d_n)を作成しt(n)と共に新規文書保持部802に保持する。そして、ステップS903に移る。

【0072】ステップS903では、その時アクセスしているフォルダが文書保持部801に保持されているフォルダの何番目であるかを示すカウンタxの値を初期値の1に設定する。そして、ステップS904に移る。

【0073】ステップS904では、カウンタxの値と文書保持部801中のフォルダの数Nを比較し、 $x \leq N$ であればステップS905に移る。 $x > N$ であれば処理を終了する。

【0074】ステップS905では、文書保持部801中のx番目のフォルダf(x)と新規文書d(n)に対してスコア $s = g(v(d_n), v(f_x))$ を計算する。そして、ステップS906に移る。

【0075】ステップS906では、ステップS905で求めたスコアsと予め設定されている閾値Scとを比較し、 $s > Sc$ であればステップS907に移る。 $s \leq Sc$ であればステップS908に移る。

【0076】ステップS907では、分類結果保持部804において、フォルダf(x)に相当する集合に新規文書d(n)を追加する。そして、ステップS908に移る。

【0077】ステップS908では、カウンタxの値を1増やす。そして、ステップS904に戻る。

【0078】また、保存先作成処理は、保存対象文書保持部805に保持された文書が、分類結果保持部804において属しているフォルダ全てを保存先保持部807に保持する。保存先変更処理は、利用者の指示に従って、保存先保持部807に保持されている保存先のリス

トにフォルダの追加もしくは削除を行なう。保存処理は、図4のフローチャートに示した先の実施例の保存処理と同様である。

【0079】本実施例の文書処理装置を用いると、文書の収集時に、各フォルダへ保存すべきと判断される文書がフォルダごとに分類されて分類結果保持部804に保持される。そこで、例えば分類結果保持部804に保持されている分類に従って、フォルダごとに文書を閲覧すれば、利用者は自らの視点に沿って関連する文書をまとめて閲覧することができる。また、保存先保持部807にデフォルトの保存先が保持されており、変更の必要がなく、変更の指示が入力手段207から入力されなければ利用者が保存先を改めて指示することなく、保存の指示のみで保存処理を行なうことも可能である。以上のように、本実施の形態の文書処理装置を用いることで、文書を視点に沿って閲覧し視点に沿った保存先が得られるので、文書の収集に要する作業が容易になる。

【0080】(他の実施例2)前記実施の形態においては、分類結果から保存先を作成し、この保存先だけを利用者が見て変更する場合について説明したが、本発明はこれに限定されるものではなく、図1に示した保存候補作成処理と組み合わせて保存候補も利用しながら保存先を変更することで収集を容易にする例について説明する。

【0081】図10は、本実施の形態に係る装置の基本構成を示すブロック図である。同図において、1001は、フォルダ及びフォルダに属する文書を保持する文書保持部である。1002は、新たに到着した文書を保持する新規文書保持部である。1003は、新規文書保持部1002に保持されている文書の保存先として適するフォルダに文書を分類する文書分類部である。1004は、文書分類部1003により分類された結果を保持する分類結果保持部である。1005は、保存処理の対象とする文書を保持する保存対象文書保持部である。1006は、分類結果保持部1004に保持されている分類結果から保存対象文書保持部1005に保持されている保存対象文書の保存先を作成する保存先作成部である。1007は、保存先作成部1006により作成された保存先を保持する保存先保持部である。1008は、保存対象文書保持部1005に保持されている文書の保存先として適するフォルダを検索し、その中から保存先保持部1007に保持されているフォルダを除いたものを保存先の候補とする保存候補作成部である。1009は、保存候補作成部1008により作成された保存先の候補を保持する保存候補保持部である。1010は、利用者により保存先保持部1007に保持されている保存先と保存候補保持部1009に保持されている保存候補を変更する保存先変更部である。1011は、保存対象文書保持部1005に保持されている文書を保存先保持部1007に保持されている保存先のフォルダに属するよう

文書保持部1001に保持する処理を行なう保存処理部である。

【0082】本実施例では、文書保持部1001、新規文書保持部1002、分類結果保持部1004、保存先保持部1007は、先に説明した図9の構成901、902、904、907と同様である。また、保存候補保持部は、先に説明した図1の構成104と同様である。各処理も先に説明した処理と同様であるが、保存先変更処理は一部異なる。本実施例における保存先変更処理では、保存先保持部1007から削除された保存先は、保存候補保持部1009に保持される。また、保存候補保持部1009に保持されている保存先候補が、保存先保持部1007に追加された場合は、それを保存候補保持部1009から削除する。

【0083】本実施例の文書処理装置を用いると、保存先の決定にあたって分類結果を修正する場合に追加するフォルダを容易に探すことができるようになり、文書の収集がより容易になる。

【0084】尚、ここまで説明した全ての実施の形態においては、保存候補検索処理及び文書分類処理において特徴ベクトルの距離関係を用いてスコアを計算する場合について説明したが、本発明はこれに限定されるものではない。文書がフォルダに属する度合いを示すスコアを得られる他の手法を用いてもよい。例えば、フォルダのデータに、利用者が与えるキーワードとその論理関係による検索条件 c を加えて、

$$f = (1, D, c, v(f))$$

とし、スコアを $s = f(c(x), d(n))$ により計算する方法や $s = f(c(x), d(n)) * C + g(v(fx), v(d_n))$ により計算する方法などがある(C は任意の定数である)。

【0085】また、フォルダ検索処理においてキーワードとその論理関係による検索条件を用いて検索する場合について説明したが、本発明はこれに限定されるものではない。フォルダを検索できる他の手法を用いてもよい。例えば、検索したいフォルダに近い別のフォルダ $f(t)$ を検索条件として、スコア s を $s = g(v(fx), v(f t))$ として計算する方法がある。あるいは、検索したフォルダの内容に近い文書 $d(t)$ を検索条件として、スコア s を $s = g(v(fx), v(d t))$ として計算する方法がある。

【0086】尚、先の実施の形態においては、利用時の検索を実現する構成としてフォルダ検索部だけを持つ場合について説明したが、本発明はこれに限定されるものではなく、文書を検索する文書検索部を設けてもよい。

【0087】尚、先の実施の形態においては、文書分類処理で全てのフォルダに対応した分類を行なう場合について説明したが、本発明はこれに限定されるものではなく、特定のフォルダに限って分類を行なってもよい。例えば、利用者が指定したフォルダだけに限定したり、一

定期間内に利用されたフォルダだけに限定したりする。

【0088】尚、先の実施の形態においては、文書分類処理でスコアを同じ方法で求めて同一の閾値と比較して判定する場合について説明したが、本発明はこれに限定されるものではなく、対応するフォルダによって、スコアを求める方法を変えてもよいし、閾値を変えてもよい。

【0089】尚、先の実施の形態においては保存候補検索処理及びフォルダ検索処理において全てのフォルダを結果として保持する場合について説明したが、本発明はこれに限定されるものではなく、一部のフォルダだけを結果として保持してもよい。例えば、スコアが予め設定された閾値を越えたものだけを保持する方法や予め設定された数や割合の範囲内で結果を保持する方法がある。

【0090】尚、先の実施の形態においては、収集時に新しいフォルダを作成しない場合について説明したが、本発明はこれに限定されるものではなく、新しいフォルダを作成して、当該フォルダを保存先保持部に追加する新規保存先作成部を設けてもよい。

【0091】尚、先の実施の形態においては、分類結果保持部に保持されている分類結果が常に保持される場合について説明したが、本発明はこれに限定されるものではなく、保存処理が行なわれたときに当該文書を分類結果から削除したり特定のフォルダに対応する分類結果を削除するなど、分類結果削除部を設けてもよい。

【0092】尚、先の実施の形態においては、フォルダ検索処理において、複数のフォルダに保存されている文書に対して関数 f の値を求める場合について説明したが、本発明はこれに限定されるものではなく、一つの文書に対して関数 f の値を1回だけ求めるような処理でもよい。例えば、1回計算した関数 f の値を保持しておく方法や、文書に対する f の値を計算した後、その値を当該文書が属するフォルダに伝播させてフォルダをスコアを合成していく方法がある。

【0093】尚、先の実施の形態においては、パターンマッチングにより関数 f の値を求める場合について説明したが、本発明はこれに限定されるものではなく、文書に対するインデックスを作成しておき、インデックスを用いて関数 f の値を求めてもよい。

【0094】

【発明の効果】以上説明したように、本発明によれば、新規文書の保存先として適したフォルダをユーザに提示できるので、ユーザの保存先選択処理が容易になる。

【0095】また、利用者の視点に沿って文書を管理できる。

【0096】また、文書の収集時或は利用時における適切なフォルダの選択が容易になる。

【0097】また、利用者の文書保持の視点が変化しても対応できる。

【0098】また、提示された保存先を変更する処理だ

けで保存先を指示できるので、保存先の指示操作が簡単になる。

【0099】また、順次新しいフォルダが作成されるので、新たに必要フォルダの作成処理が不要となり、操作性が向上する。

【0100】また、新規文書を保存するのに適したフォルダの選択を、既にフォルダに属して保持されている文書の特徴から判断するので、適切な判断が行える。

【0101】また、ユーザに報知する保存先として適したフォルダを複数用意するので、ユーザの求めるフォルダが選択され損ねる危険性が低い。

【0102】また、表示器により正確にフォルダを確認することができる。

【0103】また、検索条件に適したフォルダを提示するので、検索処理を効率よく行うことができる。

【図面の簡単な説明】

【図1】本発明に係る文書処理装置の基本構成例を示すブロック図である。

【図2】本発明に係る機能的な具体的構成を示す図である。

【図3】本発明に係る保存候補検索処理の概要を示すフローチャートである。

【図4】本発明に係る保存処理の概要を示すフローチャートである。

【図5】本発明に係るフォルダ検索処理の概要を示すフローチャートである。

【図6】本発明に係る基本構成例を示すブロック図である。

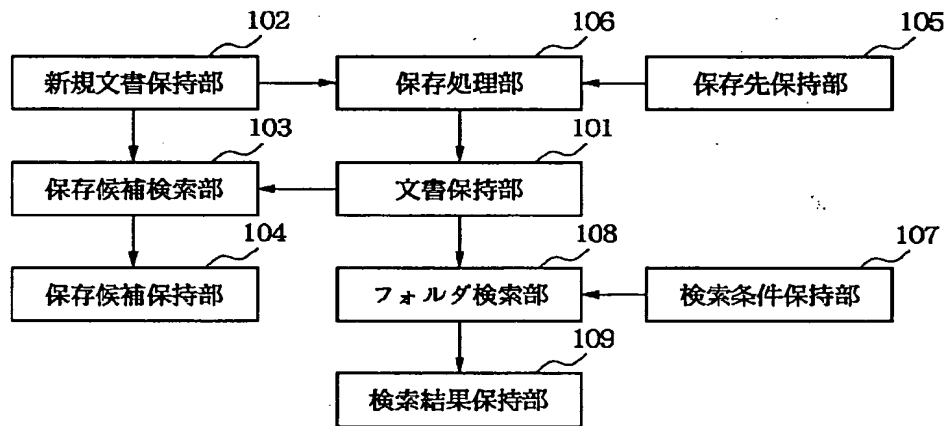
【図7】本発明に係る別の基本構成例を示すブロック図である。

【図8】本発明に係る別の基本構成例を示すブロック図である。

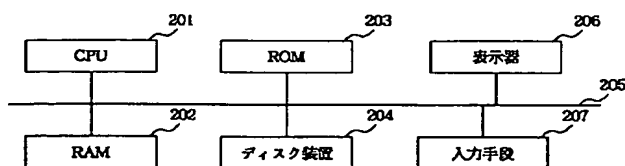
【図9】本発明に係る文書分類処理を示すフローチャートである。

【図10】本発明に係る別の基本構成例を示すブロック図である。

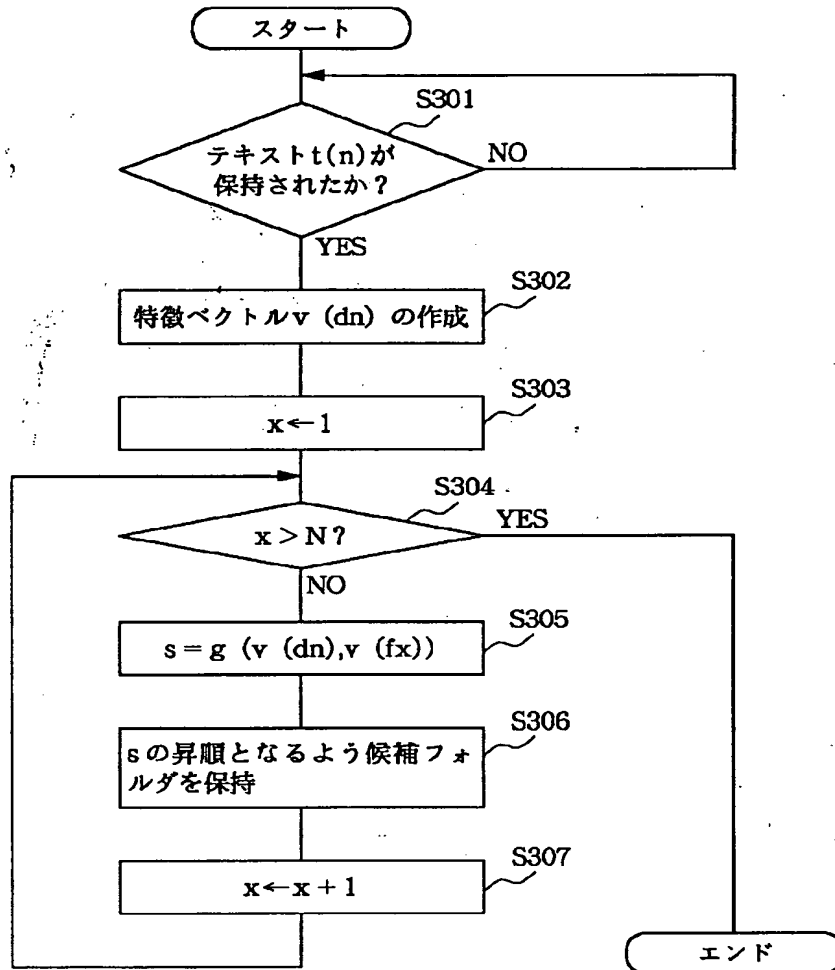
【図1】



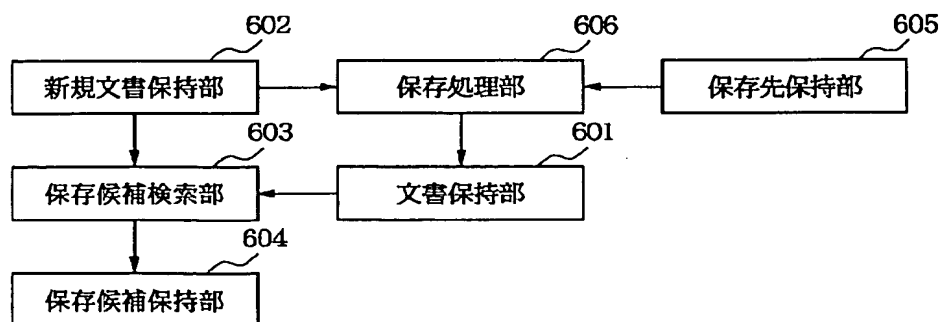
【図2】



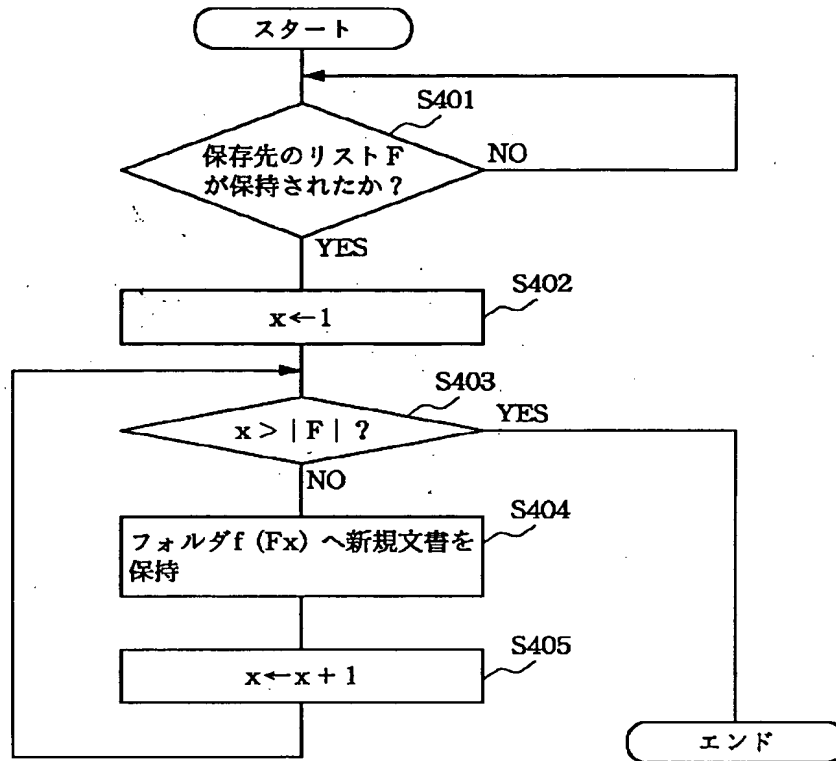
【図 3】



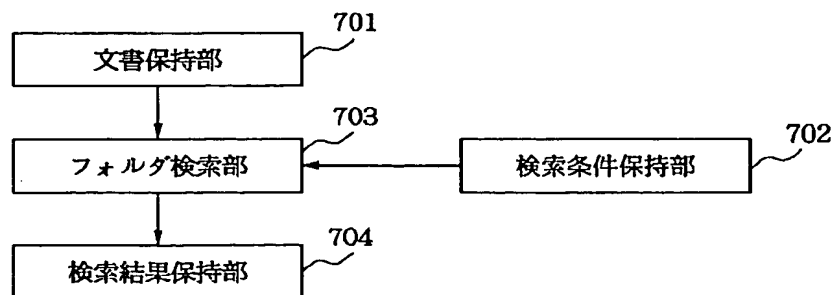
【図 6】



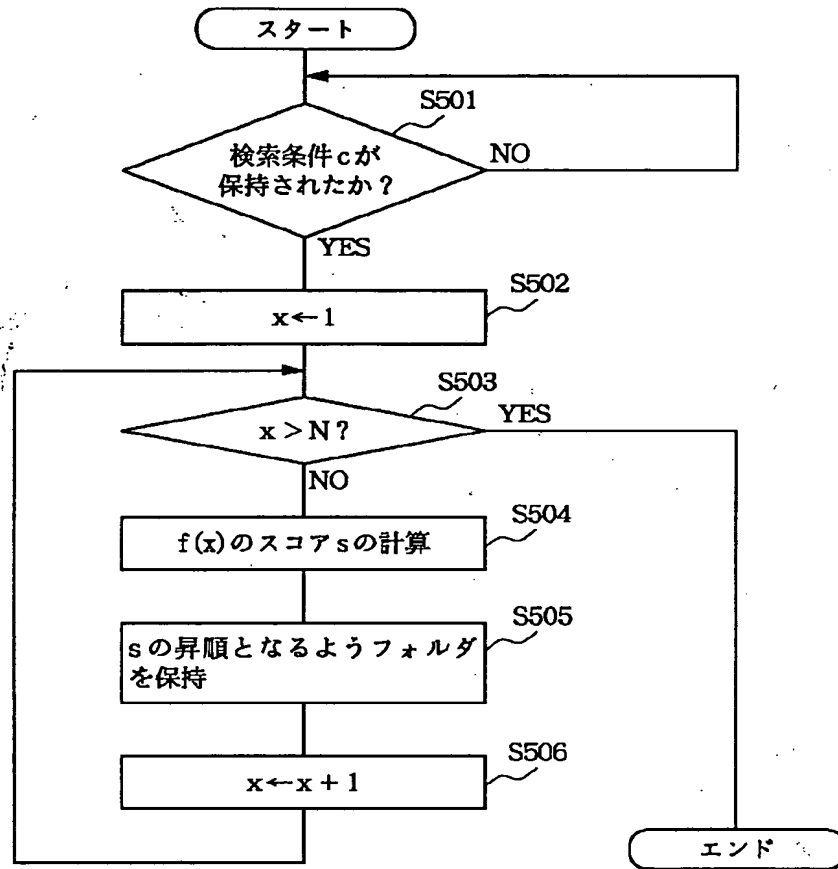
【図4】



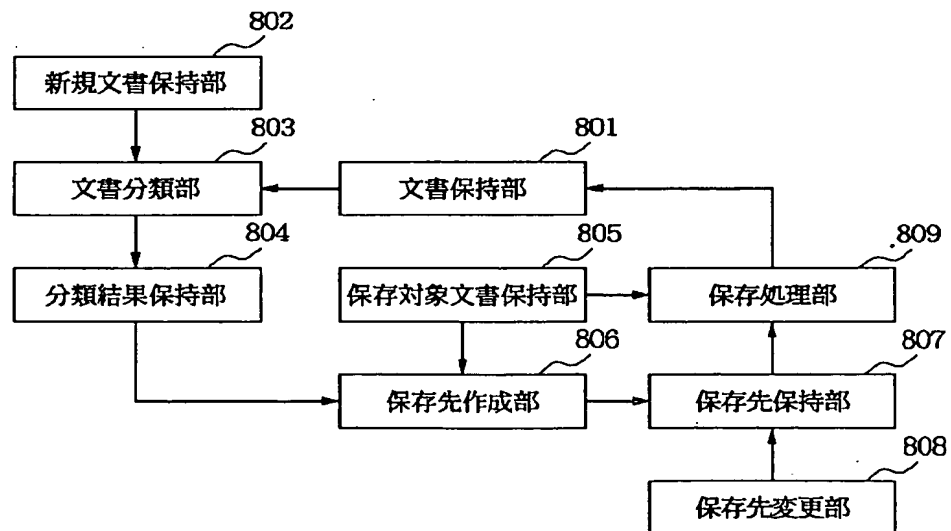
【図7】



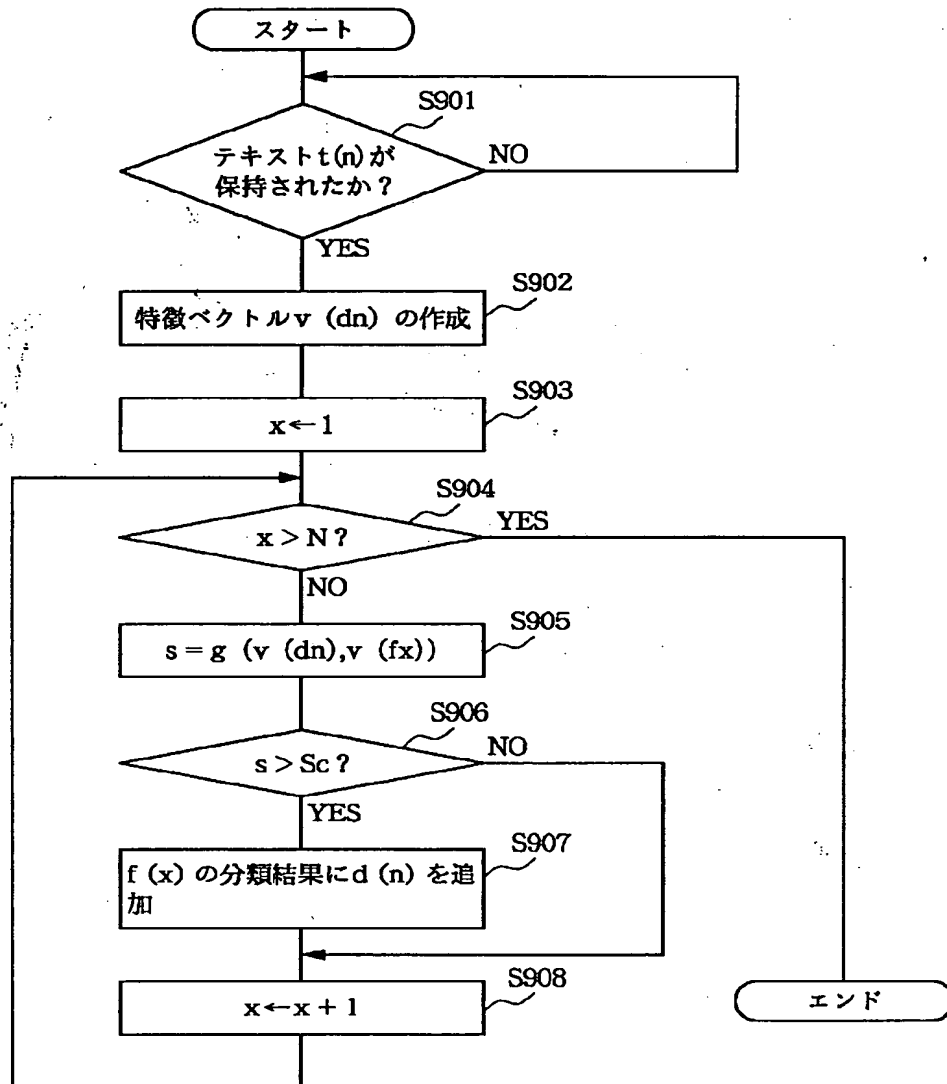
【図 5】



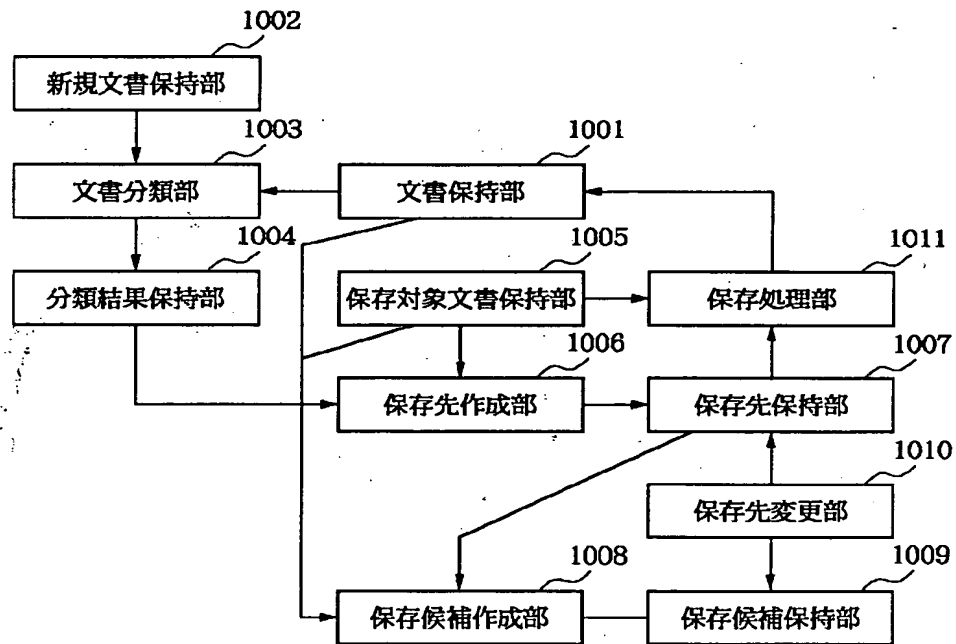
【図 8】



【図9】



【図 10】



フロントページの続き

(72)発明者 池田 裕治
東京都大田区下丸子3丁目30番2号キャノ
ン株式会社内